ES-PTAM: Event-based Stereo Parallel Tracking and Mapping

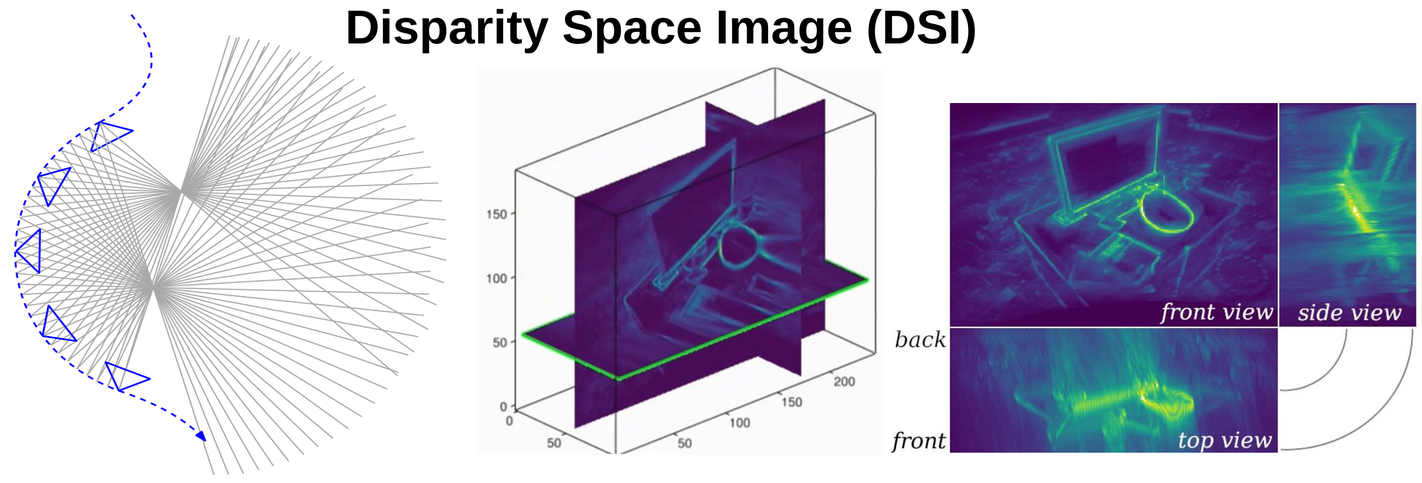
Suman Ghosh, Valentina Cavinato, Guillermo Gallego

Summary

SLAM with frame-based cameras struggles during high-speed motion and HDR lighting. We propose a SLAM front-end with stereo event cameras that

- Estimates 6-DoF camera poses and semi-dense local maps as point-clouds.
- Produces **SOTA results** (evaluated on five public datasets with diverse motions and resolutions).
- Scales efficiently in **multi-camera (≥ 2) settings**.

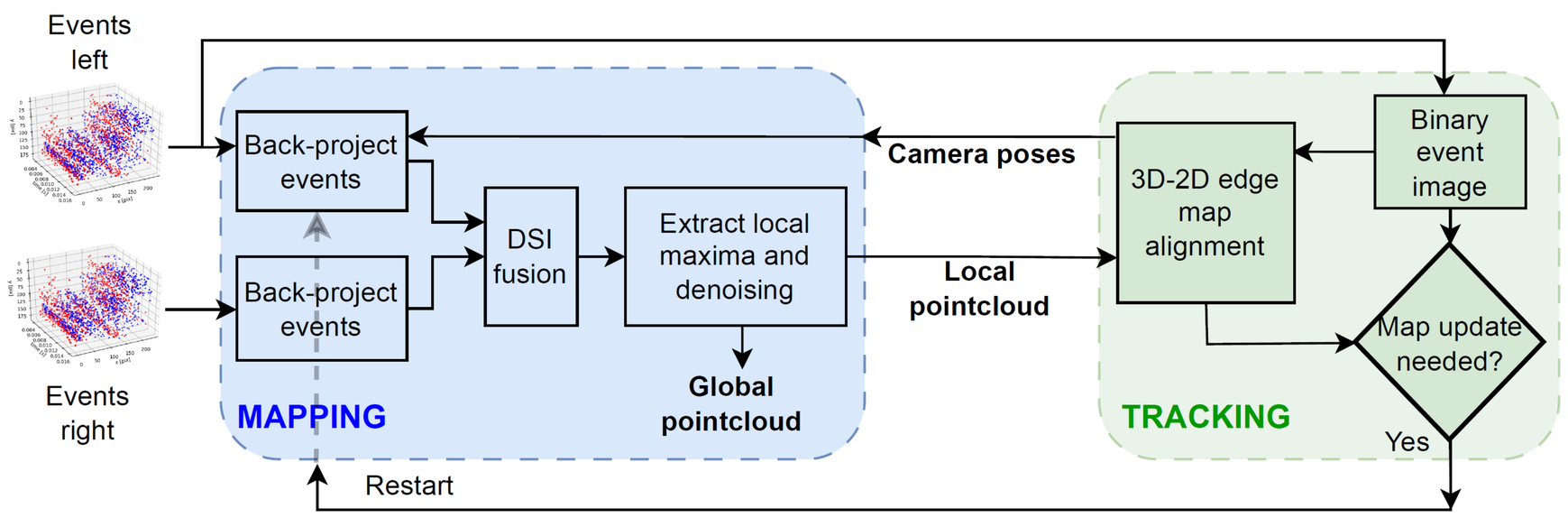
Disparity Space Image (DSI)



Visual Odometry Pipeline

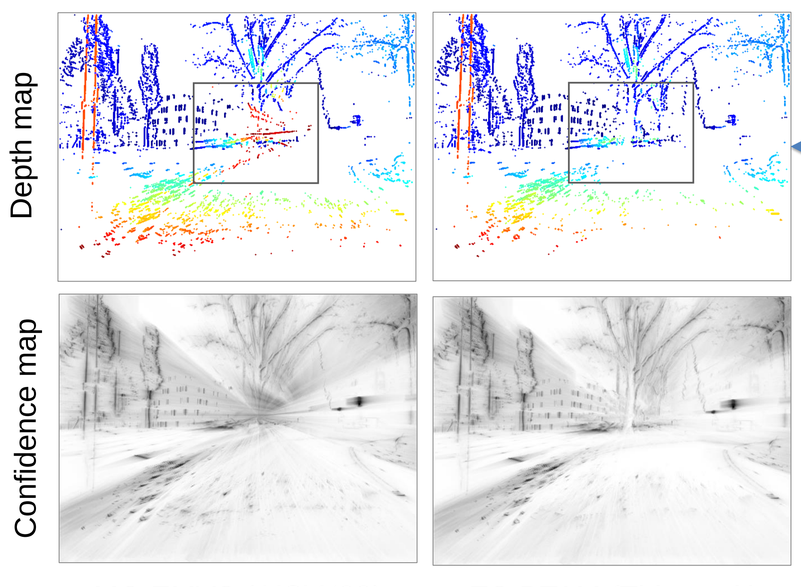
Our direct VO system combines parallelly-running modules for

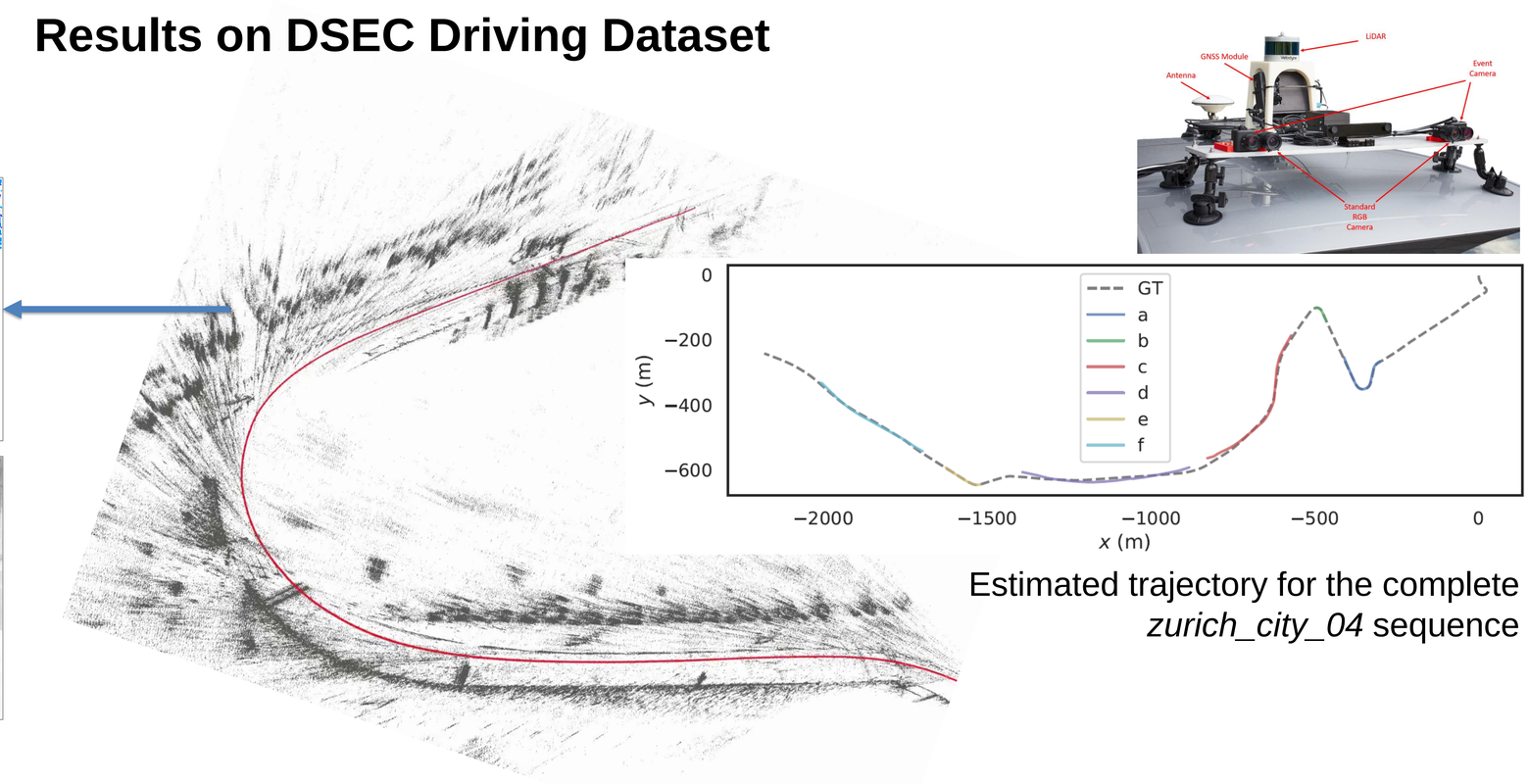
- **Mapping**: Back-project rays from event pixels using camera pose and fuse them across multiple cameras.
- **Tracking**: Minimize edge alignment error between projected point-cloud and accumulated event image.



Results on DSEC Driving Dataset

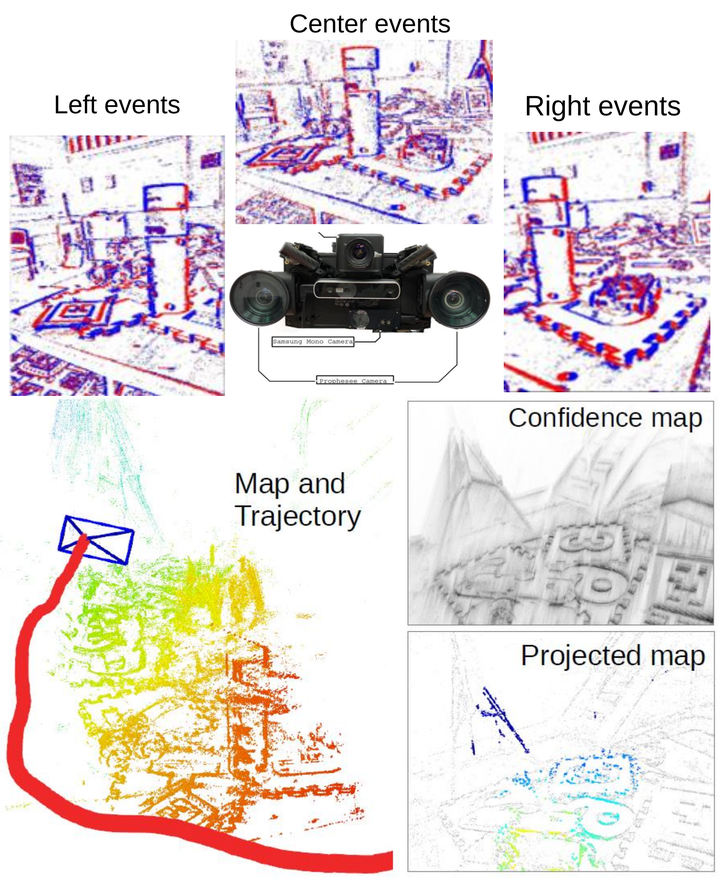
Limiting depth planes where rays are projected to improves mapping accuracy.



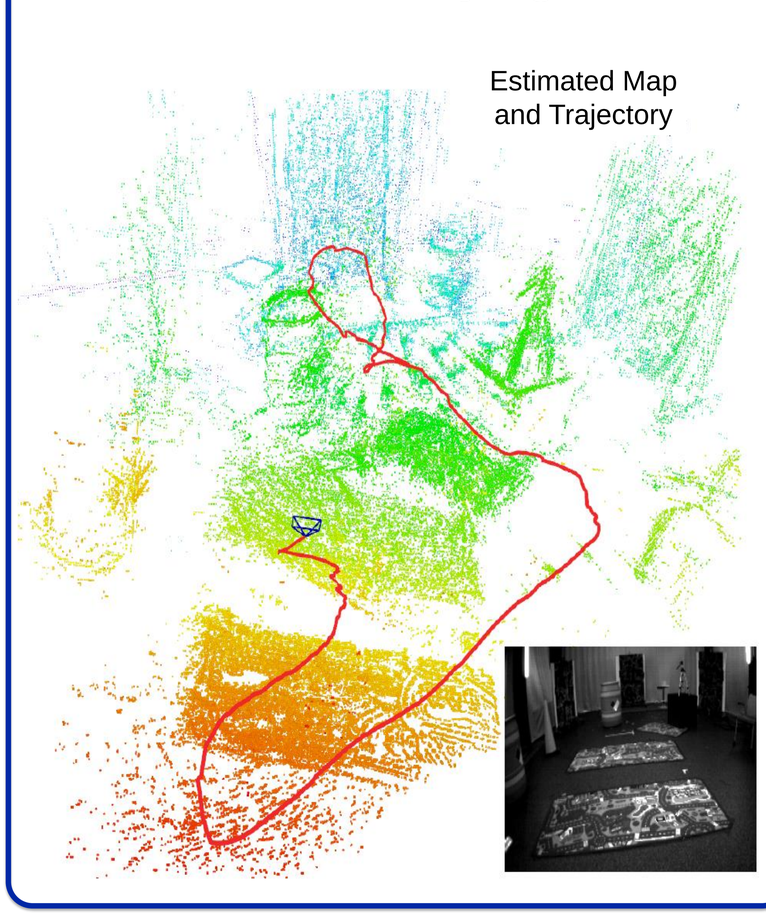


Estimated trajectory for the complete *zurich_city_04* sequence

EV-IMO2v2 Trinocular Dataset



MVSEC Drone Flying Dataset



Quantitative Results

Dataset	Sequence	ATE RMSE [cm]	
		ESVO	ES-PTAM
RPG stereo DAVIS (346 X 260)	monitor	5.8	**2.3**
	desk	7.3	**2.8**
	boxes	9.5	**4.1**
DSEC (640 X 480)	zc04a	371.1	**131.6**
	zc04b	116.6	**29.0**
	zc04c	1357.1	**1184.4**
	zc04d	2676.6	**1053.9**
	zc04e	794.9	**75.9**
TUM-VIE (1280 X 720)	1d-trans	12.3	**1.1**
	3d-trans	17.2	**8.5**
	6dof	13.0	**10.3**
	desk	12.4	**2.5**
	desk2	**4.6**	7.2

Module	Runtime on Intel Xeon(R) 4.10GHz×8 CPU
Mapper	140 ms
Tracker	6-20 ms

TUM-VIE HD Dataset

Code, Paper, Video